AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of Myophage Ec_Makalu_002, Which Infects Uropathogenic *Escherichia coli*

Gunaraj Dhungana,[a,b] Rajani Malla,[a] Manoj Rajaure,[b] Sankar Adhya[b]

[a]Central Department of Biotechnology, Tribhuvan University, Kirtipur, Nepal
[b]Laboratory of Molecular Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

ABSTRACT We isolated phage Ec_Makalu_002, which infects uropathogenic strains of *Escherichia coli*. Here, we report its complete genome sequence, annotated features, and relatedness to other phages.

One of the most common pathogens responsible for urinary tract infection (UTI) is *Escherichia coli* (1). The biggest concern about these uropathogenic strains of *E. coli* is their insensitivity to existing antibiotics and their high recurrence rates, which are linked to their ability to form both extra- and intracellular biofilm-like communities within the bladder (2, 3). To address the resistance and recurrence problem, phages are currently being suggested as an effective and alternative therapeutic (4), especially in developing countries with poor sanitation and hygiene (5). In this report, we describe the genome of Ec_Makalu_002, which was isolated from a municipal wastewater canal in Kathmandu, Nepal.

Phage Ec_Makalu_002 was originally enriched from a filtered (0.2-μm pore size) wastewater sample by infecting an aerobically growing culture of a deidentified clinical strain of uropathogenic *E. coli* at 37°C in LB broth. The host was obtained from the National Public Health Laboratory in Nepal. A spot test demonstrated that Ec_Makalu_002 also possessed the ability to propagate on a laboratory strain of *E. coli* K-12 (MG1655), which was utilized for purification using the soft-agar overlay method (6). A high-titer phage lysate (2.3×10^9 PFU/ml) was used to extract the genomic DNA using the phenol-chloroform extraction method. The DNA library was prepared using an Illumina Nextera XT kit, and whole-genome sequencing was performed on a NextSeq 500 platform, resulting in 9,387,393 150-bp paired-end reads. Reads were inspected for overall quality using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), adaptor sequence trimming was performed using Trimmomatic (7), and *de novo* sequence assembly was done using SPAdes 3.13.1 (8). All tools were run with default parameters unless otherwise specified. The largest assembled contig (164,751 bp with 4,895-fold coverage) was obtained with 77-bp identical sequences at each end, suggestive of a circularly permuted DNA packaging mechanism. The assembled genome was closed with PCR using primers (5′-GCGATTGATGCTATTCAAATGCAG-3′ and 5′-CCGATAATCTCTTTTAGACCG GACG-3′) facing off the ends and manually corrected matching of the Sanger sequencing reads. Tools available at the Galaxy and WebApollo instances via the Center for Phage Technology (CPT) (https://cpt.tamu.edu/galaxy-pub/) were used for structural and functional annotation of the assembled contig (9, 10). For example, GLIMMER 3.0 (11) and MetaGeneAnnotator 1.0 (12) were used to identify coding genes, tRNA prediction was done with ARAGORN 2.36 (13), and transcriptional terminators were manually inspected based on prediction from TransTermHP (14). Gene functions were predicted largely by similarity to the Canonical Phages database based on BLASTp searches (15) and/or confirmed using InterProScan (16) and TMHMM (17), tools that were available in the CPT WebApollo interface (https://cpt.tamu.edu/galaxy-pub/).

Citation Dhungana G, Malla R, Rajaure M, Adhya S. 2020. Complete genome sequence of myophage Ec_Makalu_002, which infects uropathogenic *Escherichia coli*. Microbiol Resour Announc 9:e01530-19. https://doi.org/10.1128/MRA.01530-19.

Editor Catherine Putonti, Loyola University Chicago

Address correspondence to Manoj Rajaure, manoj.rajaure@nih.gov.

Received 13 December 2019
Accepted 12 January 2020
Published 30 January 2020

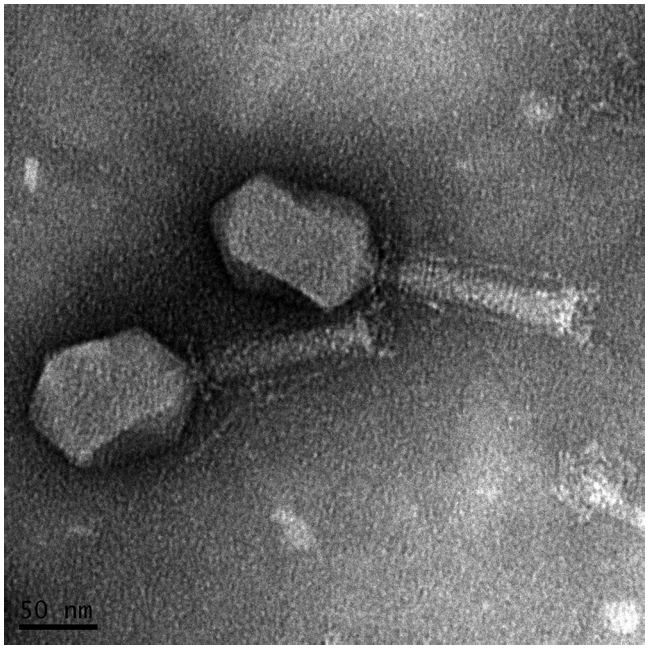

FIG 1 Transmission electron microscope image of bacteriophage Ec_Makalu_002. The phages were negatively stained with 2% uranyl acetate and observed using an FEI Tecnai T12 microscope.

The complete genome of phage Ec_Makalu_002 was 164,674 bp long with an average GC composition of 40.6%. The DNA sequence similarity of Ec_Makalu_002 was calculated using progressiveMauve 2.4.0 (18) and found to be closely related to T4-like enterobacterial phages, including ECD7 (GenBank accession number NC_041936.1; 92.26%), GEC-3S (HE978309.1; 92.23%), and Phi1 (EF437941.1; 91.43%), all of which were isolated against virulent nonlaboratory strains of *E. coli*. Consistent with the sequence analysis, imaging using transmission electron microscopy showed that Ec_Makalu_002 belongs to the *Myoviridae* family (Fig. 1). Based on its similarity to the T4-like phages and to maintain the consistency with linear genome structure in the phage database, the genome was reopened at the *rIIA* gene homolog prior to submission. This myophage encodes 274 predicted coding sequences, but no tRNA genes were detected. Putative lysis genes, holin, endolysin, and spanins were found to be scattered throughout the genome, similar to that of the T4 phage.

Data availability. The genome sequence and associated data for phage Ec_Makalu_002 were deposited under GenBank accession number MN709127, BioProject accession number PRJNA594990, and SRA accession number SRR10671636.

ACKNOWLEDGMENTS

This work was supported by funding from the Intramural Research Program of the NIH, National Cancer Institute, and part of the work was funded by the Ph.D. fellowship and research support awarded to Gunaraj Dhungana (award numbers PhD/73-74/S and T-07) by the University Grants Commission (UGC) in Nepal.

Genome sequencing was done at the CCR Genomics Core facility, and *de novo* sequence assembly was performed utilizing the computational resources of the NIH High-Performing Computation Biowulf Cluster. The phage was imaged at the facility of the National Institute of Biomedical Imaging and Bioengineering in Bethesda, MD.

REFERENCES

1. Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. 2015. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol 13:269–284. https://doi.org/10.1038/nrmicro3432.
2. Blango MG, Mulvey MA. 2010. Persistence of uropathogenic Escherichia coli in the face of multiple antibiotics. Antimicrob Agents Chemother 54:1855–1863. https://doi.org/10.1128/AAC.00014-10.

3. Mulvey MA, Schilling JD, Hultgren SJ. 2001. Establishment of a persistent Escherichia coli reservoir during the acute phase of a bladder infection. Infect Immun 69:4572–4579. https://doi.org/10.1128/IAI.69.7.4572-4579.2001.
4. Kortright KE, Chan BK, Koff JL, Turner PE. 2019. Phage therapy: a renewed approach to combat antibiotic-resistant bacteria. Cell Host Microbe 25:219–232. https://doi.org/10.1016/j.chom.2019.01.014.
5. Nagel TE, Chan BK, De Vos D, El-Shibiny A, Kang'ethe EK, Makumi A, Pirnay J-P. 2016. The developing world urgently needs phages to combat pathogenic bacteria. Front Microbiol 7:882. https://doi.org/10.3389/fmicb.2016.00882.
6. Adams MH. 1959. Bacteriophages. Interscience Publishers, Inc., New York, NY.
7. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
9. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Gruning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46:W537–W544. https://doi.org/10.1093/nar/gky379.
10. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elsik CG, Lewis SE. 2013. Web Apollo: a Web-based genomic annotation editing platform. Genome Biol 14:R93. https://doi.org/10.1186/gb-2013-14-8-r93.
11. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res 27:4636–4641. https://doi.org/10.1093/nar/27.23.4636.
12. Noguchi H, Taniguchi T, Itoh T. 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res 15:387–396. https://doi.org/10.1093/dnares/dsn027.
13. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. https://doi.org/10.1093/nar/gkh152.
14. Kingsford CL, Ayanbule K, Salzberg SL. 2007. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. Genome Biol 8:R22. https://doi.org/10.1186/gb-2007-8-2-r22.
15. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
16. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.
17. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315.
18. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147. https://doi.org/10.1371/journal.pone.0011147.